ED 076 641 TM 002 646

AUTHOR Hill, Richard K., Jr.
TITLE Estimating Total-Test Score Distributions Through Item Sampling--A New Theoretical Approach.
PUB DATE 73
NOTE 12p.; Paper presented at American Educational Research Association Meeting (New Orleans, Louisiana, February 25-March 1, 1973)

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS *Item Sampling; *Mathematical Models; *Norms; Speeches; *Statistical Analysis; *Test Results

ABSTRACT

A model for multiple choice test-taking behav.or is proposed which is different from those presently used for item sampling theory. A new theory is developed, which includes a new concept to facilitate comprehension of item sampling theory, a "number-known" score distribution. A major advantage of the model is that it accommodates data from multidimensional tests. The relationships between the proposed model and established theory are shown. The advantages and disadvantages of the proposed theory are compared to those of other techniques currently in use. (Author)

ESTIMATING TOTAL-TEST SCORE DISTRIBUTIONS THROUGH[1] ITEM SAMPLING--A NEW THEORETICAL APPROACH

Richard K. Hill, Jr.
Virginia Polytechnic Institute and State University

Matrix sampling is a sampling scheme in which samples of test items are administered to samples of subjects. Inference can then be made to a population of subjects, a population of items, or both.

Frequently, the problem of interest is the estimation of what the distribution of scores would be like had all the subjects taken all the items. The given information, of course, is the distributions of scores on the item samples by the samples of subjects. At present, there are three methods of estimating norms from item samples. This paper will present a fourth. This latest technique has a different theoretical framework and thus has advantages and disadvantages different from those now available.

Collecting data using matrix sampling. A test containing K items is divided into t subtests, each subtest containing k items. The subtests are formed by randomly selecting the items from the total test without replacement. Each of the subtests is then administered to n subjects. Although other sampling plans can be used (see Shoemaker, 1971 for a more complete exposition of the advantages and applications of matrix sampling techniques), this will be the plan used for the development of the model presented in this paper.

Currently available techniques. Three methods of estimating total-test score distributions from matrix sample results are currently available. The first, due to Keats and Lord (1962), uses the negative hypergeometric distribution as a model for total-test scores. The sufficient parameters are total-test mean and variance, and K. Previous work by Lord (1960) provides the necessary calculations to fit the negative hypergeometric. The negative hypergeometric has been used

[1] Presented at the 1973 AERA meeting, New Orleans.

successfully many times to estimate total-test score distributions. However, the assumptions underlying the model are very restrictive and it is not difficult to find examples of tests which would violate the assumptions to the extent that one would have little confidence in using results based upon the negative hypergeometric.

A second method for estimating total-test scores is due to Kleinke (1969). This technique predicts a total-test score for each individual using linear regression. The combination of all estimated total-test scores yields the estimated total-test score distribution. A major disadvantage of this technique is the severe jaggedness of the estimated distributions.

A third method for estimating the total-test score distributions is the empirical Bayes' estimation technique, used by Lord (1969). This method uses an empirical Bayesian procedure to obtain minimum squared error estimators for total-test score parameters. Its major disadvantage is the requirement of large amounts of data to avoid uninterpretable results. Also, no unique solution can be found to any problem unless one makes some further assumptions.

The general framework of the proposed model. The models noted above, either directly or by implication, make use of a true-score model. That is, given a distribution of scores on an item sample, one does not make a direct estimate of the distribution of scores on the total-test. Since scores from both distributions contain error, it is first necessary to estimate some error-free distribution based on the item sample results, use that estimate to predict the shape of the error-free distribution on the total-test, and then add the error back in to obtain an estimated total-test score distribution.

As noted above, the error-free distribution used in the past has been the true-score distribution. The model proposed by this paper utilizes what has been called a "guessing-free" distribution as the intermediate step. A guessing-free distribution is the distribution of scores obtained when no one guesses at any item, but answers only those questions of which he is sure of the answer.

The model is used in the construction of an intermediate step in much the same way that the true-score models were used in earlier methods. Thus, the observed scores on the item sample will be used to estimate the guessing-free distribution on the item sample. The second step is to use these results to estimate the guessing-free distribution on the full test. This is then used to estimate what the total-test score distribution would have looked like if all subjects had taken all the items, rather than just a sample of them.

<u>The</u> <u>proposed</u> <u>model</u>. As a beginning, to test whether there was any future in using guessing-free distributions to estimate total-test results, the simplest possible model for estimating the guessing-free distributions was used. The model simply views an individual's correct response on an item as having possibly occurred by one of only two ways--he knew the answer, and thus chose the correct response with certainty, or he did not know it, and randomly chose an alternative. Another simplification is the assumption of population data on each sample. Under these assumptions, an observed score distribution is simply the combination of two simpler distributions: the guessing-free distribution and the distribution of guesses.

The distribution of guesses, under the stated assumptions, is a binomial on the remaining items. Thus, if f(x) is the distribution of observed scores and g(x) is the distribution of guessing-free scores,

$$f(x) = \sum_{i=0}^{x} \binom{k-i}{x-i} p^{x-i} q^{k-x} g(i), \qquad (1)$$

where $p = 1/\#$ of options per question, and
$q = 1 - p$.

From (1) it can be shown (Hill, 1972, pg. 113) that

$$g(x) = \sum_{i=0}^{x} \binom{k-i}{x-i} (-p)^{x-i} q^{i-k} f(i) \qquad (2)$$

This allows us now to calculate a guessing-free distribution from an observed-score distribution and vice versa. The remaining step is to calculate the guessing-free distribution on the full test given the guessing-free distribution on the item sample.

This is clearly a hypergeometric problem, since items from the full test are randomly sampled without replacement. Thus, if an individual has a guessing-free score of a on the full test, the probability that his guessing-free score on the item sample will be b is:

$$P(b|a) = \frac{\binom{a}{b}\binom{kt-a}{k-b}}{\binom{kt}{k}} \quad \text{if } b < a < k(t-1) + b \qquad (3)$$

$$= 0 \qquad \text{otherwise}$$

If h(x) is the distribution of guessing-free scores on the full test, then

$$g(x) = \sum_{i=0}^{kt} P(x|i)\ h(i) \qquad (4)$$

Unfortunately, no unique solution for h(x) exists (Hill, 1972, pg. 28). However, it is possible to calculate the first k moments of the guessing-free distribution on the full test, given the first k moments of the guessing-free distribution on the item sample. The method for doing so will not be shown here, but is available from Hill (1972, pp. 25-30).

The remaining problem, then, is to find what problems exist in attempting to define a distribution that ranges over kt+1 points when only k moments are known. Although no unique solution exists, it is possible that the family of possible solutions is so similar that the errors of sampling would outweigh the problem of non-uniqueness, and the selection of any arbitrary distribution from the family of possible solutions would yield satisfactory results.

Defining maximally different distributions. Since there is no unique solution, the question arises as to how different any two possible solutions could be, since if the differences are small, they might be insignificant compared to other problems likely to be encountered in item sampling. It will be necessary to define operationally the term "maximally different."

One possible definition entails the use of the Kolmogorov-Smirnov *D* statistic, which is the maximum difference between two cumulative frequency distributions. This statistic was not used in the definition, because no way was discovered to maximize *D* between two distributions. Other possibilities included maximizing χ^2 or some similar statistic, but trying to maximize a statistic by changing two distributions simultaneously has proved intractable. Since the problem of non-uniqueness can be considered to be caused by the lack of well-defined higher-order moments, it seems plausible to use as a criterion of maximal difference the maximally different higher-order moments. At first, it seems as though a problem might be encountered in deciding which moment to use for the definition; i.e., if the first five moments are uniquely determined, should the two maximally different solutions be defined as those which maximize the differences in the sixth moment, the seventh moment, or the *kt*th moment? It can be demonstrated that it makes no difference which moment is chosen, because the solution remains the same.

Maximizing the difference between undefined higher-order moments. Maximizing the difference between a moment of two distributions is not difficult because one can find the distribution that maximizes the moment, another that minimizes it, and have the solution. It is much easier to operate this way, for by solving for one distribution at a time (rather than by finding both simultaneously, as would be necessary with the Kolmogorov-Smirnov *D*), one can use linear programming. The following linear constraints can be set up from the estimated moments which have been calculated:

$$\sum_{i=0}^{kt} i^r\,h(i) = E(h^r) \qquad 0 \le r \le k$$

The following becomes the objective function:

$$\sum_{i=0}^{mn} i^p\,h(i), \text{ where } p \text{ is the moment to be maximized.}$$

The objective function, of course, is simply the *p*th moment around the origin. It can ~~be both~~ maximized and minimized through linear programming, yielding the two sets of number-known score distributions that are, by the definition adopted above, maximally different.

The first question which comes to mind is deciding which moment to maximize and minimize. The answer is that it does not matter. The same solution will be obtained no matter which moment is chosen. The proof for this is presented in Appendix B of Hill (1972).

Convergence of maximum and minimum distributions. It was intended to use linear programming in the following manner to arrive at a solution:

1. Find the distributions that will maximize and minimize the *k+1*th moment, given the constraints of the first *k* moments.

2. Under the constraint of the boundaries defined by the two distributions calculated in step 1, maximize and minimize the *k+2*th moment.

3. Under the constraints of the new boundaries calculated in step 2, maximize and minimize the *k+3*th moment.

4. Continue operating until the differences between the maximizing and minimizing distributions are very small, or the iterative process is done for the *kt*th moment.

There is a flaw in the proposed solution, however, since given only the first *k* moments of a distribution over *kt+1* points, there should be no unique solution. But if this approach worked, there would be.

Of course, this approach failed, because the functions do not converge. Once step 1 has been completed, further calculations are meaningless, for the same solution sets are obtained every time. For this reason, the following approach was proposed to arrive at a final solution:

1. Find the distributions that will maximize and minimize the k+1th moment, given the constraints of the first k.
2. Calculate the mean of the minimum and maximum values of the k+1th moment, and use this number as the value of it.
3. Calculate the minimum and maximum of the k+2th moment, under the constraint of the first k moments and the mean calculated in step 2.
4. Calculate the mean of these two moments, and use them as another constraint to minimize and maximize the k+3th moment.
5. Continue in this manner until the differences between the minimum and maximum distributions obtained are very small, or the iterative process is done for the ktth moment.

This solution system is not preferable to the first, for it adds something which is not there. It assumes that the undefined higher-order moments are the means of their constrained minimum and maximum possible values. It does have a major advantage over other possible solutions, however, in that a solution is always obtained. It will be shown, moreover, that the solution sets are fairly well defined after the first k moments are calculated, if k is any substantial size at all. This further work acts as a refinement only. It is an attempt to arrive at a unique solution, rather than have two similar distributions to report.

Lack of solutions due to sampling error. Preliminary investigation of the approach which was outlined above yielded some results which in one sense were very disappointing, yet in another sense, demonstrated the similarity of any two solutions very well. The problem encountered involved the lack of solutions due to sampling error. An example will help to clarify that statement.

Data were collected on a 25 item test, which had been divided into 5 item samples of 5 items each. Each item had five options. Each of the samples was administered to approximately 130 people. The results for one sample were as follows:

X	f	p
0	5	.0394
1	10	.0787
2	26	.2047
3	19	.1496
4	32	.2520
5	35	.2756

The estimated guessing-free score distribution was:

X	p
0	.1201
1	.0721
2	.2941
3	.0304
4	.2642
5	.2190

The estimated total-test mean and variance were 16.61 and 35.86, respectively, which are close to the criterion total-test values of 17.08 and 29.03, obtained from a sample of 602 subjects.

The first five moments of the item sample guessing-free known score distribution were as follows:

Moment	Value
1	2.9035
2	11.2254
3	47.5344
4	211.7729
5	971.8896

These yielded the following estimated moments for the guessing-free score distribution on the total-test:

Moment	Value
1	14.5177
2	264.1731
3	5286.5078
4	111439.8125
5	2338562.0000

The values seem reasonable. However, there is no distribution which can be constructed for a 25 item test that has these 5 moments. Given the first four moments, it can be shown, using linear programming, that the value for the fifth moment must lie between 2,409,923 and 2,441,843.

As pointed out above, this result is both encouraging and discouraging. First, it is discouraging because no solution can be found. Second, however, it is very encouraging, because it implies that the random sampling fluctuation in the fifth moment is greater than any error in the theory due to non-uniqueness of a solution. That is, this result indicates that the difference between any two solutions we might arrive at due to being able to estimate only the first five moments are less than the differences we might expect to find from random sampling fluctuations. This gives confidence to the original assumption that any way of getting the final solution after fixing the first n moments is credible. The error in the difference between any two plausible solutions is probably substantially less than the sampling error involved.

To avoid the problem of estimating impossible moments, the method of arriving at a final estimated number-known score distribution has been revised as follows:

1. Each estimated moment is checked, in order, to see if it is a possible value, given the values of the lower-order moments. If all estimated moments are plausible, go to step 3. Upon discovering an impossible value, go to step 2.
2. The moment estimated by the equations lies outside the minimum and maximum possible values, given the estimated lower-order moments. Therefore, discard the estimated value and use the mean of the maximum and minimum possible values.
3. Calculate the minimum and maximum value of the next higher order moment, and use the mean of these two values to be the estimate for that moment.
4. Calculate the values for the next high-order moment, as in step 3. Continue until the values are close to each other, or until the ktth moment has been estimated.

This technique insures that a solution will be found. Some of the moments of the solution may not be equal to the estimated moments, but this will be true only if the higher order estimated moments have been shown to have impossible values, due to sampling error.

We now have developed all the steps necessary to estimate a total-test score distribution given the observed scores on an item sample. Using equation (2), it is possible to estimate the guessing-free distribution on the item sample. From these results, we can use the procedure outlined above to estimate the guessing-free distribution on the full test. This result yields an estimated total-test score distribution from equation (1).

The results of this model to date have been encouraging. One study has shown it to be at least as effective as either the hypergeometric or linear prediction in estimating total-test score distributions (Hill, 1972). It is anticipated that the use of more sophisticated techniques in estimating the guessing-free distribution will reduce the errors of prediction.

REFERENCES

Hill, R. K. An alternative model for estimating total-test score distributions following item sampling. Unpublished Ph.D. dissertation, Syracuse University, Syracuse, New York, 1972.

Keats, F. A. and Lord, F. M. A theoretical distribution for mental test scores. *Psychometrika*, 1962, *27*, 59-72.

Kleinke, D. J. A linear-prediction approach to developing test norms based on item-sampling. Unpublished Ed.D. dissertation, State University of New York at Albany, 1969.

Lord, F. M. Use of true-score theory to predict moments of univariate and bivariate observed-score distributions. *Psychometrika*, 1960, *25*, 325-342.

Lord, F. M. Estimating true-score distributions in psychological testing (an empirical Bayes estimation problem). *Psychometrika*, 1969, *34*, 259-299.

Shoemaker, D. M. Principles and procedures of multiple matrix sampling. Southwest Regional Laboratory, Inglewood, California, 1971.